Lecture 3

Introduction to optimality conditions

Magnus Önnheim
2012-11-05

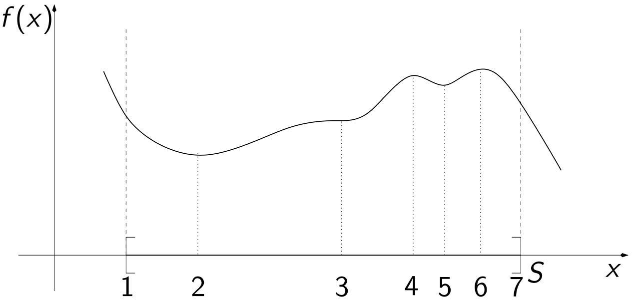
Introduction

Existence of optimal solutions

Optimality conditions
- Unconstrained case
- Convex feasible set

$$\text{minimize } f(\mathbf{x}), \quad (1a)$$
$$\text{subject to } \mathbf{x} \in S, \quad (1b)$$

$S \subseteq \mathbb{R}^n$ nonempty set, $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ a given function



(i) **boundary points** of S

(ii) **stationary points**, that is, where $f'(x) = 0$

(iii) **discontinuities** in f or f'

Here:

(i) 1, 7

(ii) 2, 3, 4, 5, 6

(iii) none

- There must exist an optimal solution, **Existence condition**
- No other point than 1–7 can be optimal **Necessary condition**
- Thus, one of the points 1–7 **is** optimal. By direct computation, it must be 2 **Sufficient condition**

- $\mathbf{x}^* \in S$ is a **global minimum** of f over S if it attains the lowest value of f over S:

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \qquad \mathbf{x} \in S$$

- $\mathbf{x}^* \in S$ is a **local minimum** of f over S if there exists a small enough ball intersected with S around $\mathbf{x}^*$ such that it is an optimal solution in that smaller set: with $B_\varepsilon(\mathbf{x}^*) := \{\, \mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{y} - \mathbf{x}^*\| < \varepsilon \,\}$ being the Euclidean ball with radius ε centered at $\mathbf{x}^*$, we get

$$\exists \varepsilon > 0 \text{ such that } f(\mathbf{x}^*) \leq f(\mathbf{x}), \qquad \mathbf{x} \in S \cap B_\varepsilon(\mathbf{x}^*)$$

- $\mathbf{x}^* \in S$ is a **strict local minimum** of f over S if $f(\mathbf{x}^*) < f(\mathbf{x})$ holds above for $\mathbf{x} \neq \mathbf{x}^*$

Introduction

Existence of optimal solutions

Optimality conditions
- Unconstrained case
- Convex feasible set

- We call S **open** if for every $\mathbf{x} \in S$ there is an $\varepsilon > 0$ such that $\{z \in \mathbb{R}^n \mid \|z - x\| < \varepsilon\} = B_\varepsilon(\mathbf{x}) \subset S$. Ex: the interval $(0, \infty)$

- We call S **closed** if $\mathbb{R}^n \backslash S$ is open. Ex: the interval $[0, \infty)$.

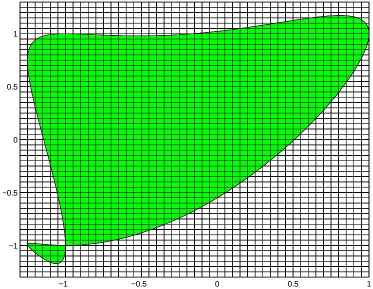
- A **limit point** of a set S is a point $\mathbf{x}$ such that there is a sequence $\{\mathbf{x}_k\}_{k=1}^{\infty} \subset S$ where $\mathbf{x}_k \to \mathbf{x}$.

- Can equivalently define a closed as a set which contains all its limit points.

- If S is both closed and bounded it is called **compact**.

Proposition: (Convergent subsequences)
Let S be compact and $\{\mathbf{x}_k\}_{k=1}^{\infty} \subset S$. Then there is a **convergent subsequence** $\{\mathbf{x}_{k_i}\}_{i=1}^{\infty}$ where $\lim_{i\to\infty} \mathbf{x}_{k_i} = \mathbf{x}$ for some $\mathbf{x} \in S$.

Intuition:



- A sequence is an infinite set of points in S
- Divide S into small blocks.
- At least one box contains infinitely many points in the sequence.

Weierstrass' Theorem (simple version): Consider the problem (1), where S is a nonempty, closed and bounded set and f is continuous on S. Then, there exists a nonempty, closed and bounded (i.e., compact) set of optimal solutions to the problem (1)

Proof.

- Construct a problem such that no optimal solution exists.
- How does it violate Weierstrass theorem?

2 min.

$S \subseteq \mathbb{R}^n$ nonempty and closed, $f : S \to \mathbb{R}$

- f is **weakly coercive** with respect to the set S if either S is bounded or
$$\lim_{\substack{\|\mathbf{x}\| \to \infty \\ \mathbf{x} \in S}} f(\mathbf{x}) = \infty$$
holds
- The weak coercivity of $f : S \to \mathbb{R}$ is equivalent to the property that f has bounded level sets (Why?)

$S \subseteq \mathbb{R}^n$ nonempty and closed, $f : S \to \mathbb{R}$

- f is **lower semi-continuous** at $\bar{\mathbf{x}} \in S$ if the value $f(\bar{\mathbf{x}})$ is less than or equal to every limit of f as $\mathbf{x}_k \to \bar{\mathbf{x}}$
  In other words, f is lower semi-continuous at $\bar{\mathbf{x}} \in S$ if

$$\mathbf{x}_k \to \bar{\mathbf{x}} \qquad \Longrightarrow \qquad f(\bar{\mathbf{x}}) \leq \liminf_{k\to\infty} f(\mathbf{x}_k)$$

- Lower semi-continuity of f is equivalent to the closedness of all its sub-level sets $\mathrm{lev}_f(b) = \{\mathbf{x} \in S \mid f(\mathbf{x}) \leq b\}$, $b \in \mathbb{R}$, as well as the closedness of its epigraph $\{(\mathbf{x}, y) : \mathbf{x} \in S, y \geq f(\mathbf{x})\}$ (Why?)
- Lower semi-continuous functions in one variable have the appearance shown in the figure on the next slide.
- **Continuity** $\Leftrightarrow$ f and $-f$ lower semi-continuous.

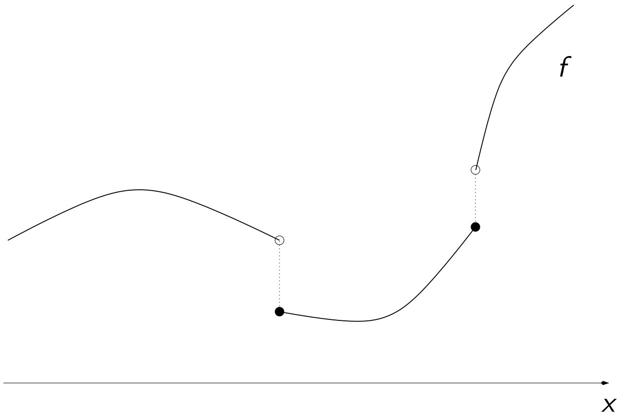


Figure: A lower semi-continuous function in one variable

Weierstrass' Theorem: Let $S \subseteq \mathbb{R}^n$ be a nonempty and closed set, and $f : S \to \mathbb{R}$ be a lower semi-continuous function on S. If f is weakly coercive with respect to S, then there exists a nonempty, closed and bounded (thus compact) set of optimal solutions to the problem (1)

Proof. Let $\bar{\mathbf{x}} \in S$ be any feasible solution. The level set $\bar{S} = \{\mathbf{x} \in S \mid f(\mathbf{x}) \leq f(\bar{\mathbf{x}})\}$ is closed (by lower semi-continuity) and bounded (by weak coercivity). Clearly the optimal solutions of (1) and $\min_{\mathbf{x} \in \bar{S}} f(\mathbf{x})$ coincide. The theorem follows in the same way as the simple version.

Example: $f(\mathbf{x}) = \|\mathbf{x} - \mathbf{y}\|$, S any closed non-empty set, any $\mathbf{y} \in \mathbb{R}^n$. Weierstrass Theorem yields that for any point $\mathbf{y} \in \mathbb{R}^n$, there is always (at least) one 'nearest' point in S.

Introduction

Existence of optimal solutions

Optimality conditions

- Unconstrained case
- Convex feasible set

Consider the problem (1), where S is a convex set and f is convex on S. Then, every local minimum of f over S is also a global minimum

Proof.

Intuitive image: If $\mathbf{x}^*$ is a local minimum, then f cannot go down-hill from $\mathbf{x}^*$ in any direction, but if $\bar{\mathbf{x}}$ has a lower value, then f has to go down-hill sooner or later. This cannot be the shape of any convex function.

We call $f : \mathbb{R}^n \to \mathbb{R}$ **differentiable** at $\mathbf{x}_0$, if there is a $\nabla f(\mathbf{x}_0)$ such that

$$f'(\mathbf{x}_0; \mathbf{d}) = \lim_{t \to 0} \frac{f(\mathbf{x}_0 + t\mathbf{d}) - f(\mathbf{x}_0)}{t} = \nabla f(\mathbf{x}_0)^{\mathrm{T}} \mathbf{d}$$

for all $\mathbf{d}$, and call f differentiable if this holds for all $\mathbf{x}_0$. If f is differentiable at $\mathbf{x}_0$, Taylor's Theorem guarantees the existence of a 'function' $o(t)$ such that

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^{\mathrm{T}}(\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|),$$

$$\lim_{t \to 0^+} \frac{o(t)}{t} = 0.$$

If $\nabla f(\mathbf{x})$ is continous, then f is called continuously differentiable, denoted by $f \in C^1$. Similarly defined is $f \in C^2$, where we also have

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^{\mathrm{T}}(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^{\mathrm{T}} \nabla^2 f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|^2),$$

where the matrix $\nabla^2 f(\mathbf{x}_0) \in \mathbb{R}^{n \times n}$ is called the **Hessian** of f at $\mathbf{x}_0$.

If $\mathbf{x}^*$ *is a local minimum of* f *on* $\mathbb{R}^n$ and $f \in C^1$, then $\nabla f(\mathbf{x}^*) = \mathbf{0}$

Proof.

Note: the other direction is not true, this is a **necessary** condition (consider $f(x) = x^3$).

Descent direction: Let $\mathbf{x} \in \mathbb{R}^n$. We call $\mathbf{p}$ a **direction of descent** with respect to f at $\mathbf{x}$ if

$$\exists \delta > 0 \text{ such that } f(\mathbf{x} + \alpha \mathbf{p}) < f(\mathbf{x}) \text{ for every } \alpha \in (0, \delta]$$

Suppose that $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is in C^1 around a point $\mathbf{x}$ for which $f(\mathbf{x}) < +\infty$, and that $\mathbf{p} \in \mathbb{R}^n$. If $\nabla f(\mathbf{x})^{\mathrm{T}}\mathbf{p} < 0$ then the vector $\mathbf{p}$ defines a direction of descent with respect to f at $\mathbf{x}$.

$$\mathbf{x}^* \textit{ is a local min of } f \textit{ on } \mathbb{R}^n \implies \begin{cases} \nabla f(\mathbf{x}^*) = \mathbf{0}^n; \\ \nabla^2 f(\mathbf{x}^*) \ \textit{ is p.s.d.} \end{cases}$$

Proof.

$$\left.\begin{array}{r} \nabla f(\mathbf{x}^*) = \mathbf{0}^n \\ \nabla^2 f(\mathbf{x}^*) \ \textit{ is p.d.} \end{array}\right\} \implies \mathbf{x}^* \ \textit{ is a strict local min of } f \ \textit{ on } \mathbb{R}^n$$

Proof.

- *Note:* $n = 1$: $x^* \in \mathbb{R}$ is a local minimum $\implies f'(x^*) = 0$ and $f''(x^*) \geq 0$
- *Note:* $n = 1$: $f'(x^*) = 0$ and $f''(x^*) > 0 \implies x^* \in \mathbb{R}$ is a strict local minimum

Let $f \in C^1$, and f be convex. Then,

$$\mathbf{x}^* \text{ is a global minimum of } f \text{ on } \mathbb{R}^n \quad \Longleftrightarrow \quad \nabla f(\mathbf{x}^*) = \mathbf{0}^n$$

Proof.

Suppose $S \subseteq \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is in C^1 around $\mathbf{x} \in S$

(a) If $\mathbf{x}^* \in S$ is a local minimum of f on S then $\nabla f(\mathbf{x}^*)^{\mathrm{T}}\mathbf{p} \geq 0$ holds for every feasible direction $\mathbf{p}$ at $\mathbf{x}^*$

(b) Suppose that S is convex and that f is in C^1 on S. If $\mathbf{x}^* \in S$ is a local minimum of f on S then

$$\nabla f(\mathbf{x}^*)^{\mathrm{T}}(\mathbf{x} - \mathbf{x}^*) \geq 0, \qquad \mathbf{x} \in S \tag{2}$$

Proof.

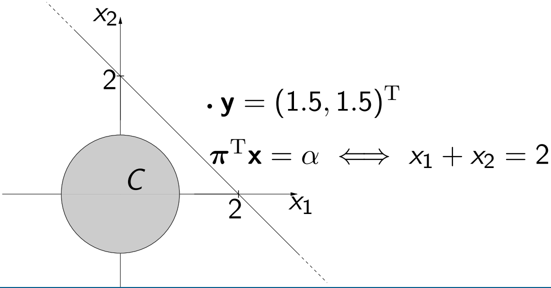
- We refer to (2) as a **variational inequality** and $\mathbf{x}^*$ as **stationary**
- *Suppose* $S \subseteq \mathbb{R}^n$ *is nonempty and convex. Let* $f \in C^1$ *on* S*, convex. Then,*

$$\mathbf{x}^* \text{ is a global minimum of } f \text{ on } S \iff (2) \text{ holds}$$

- *Proof.*
- Compare with the case $S = \mathbb{R}^n$!

- The **Separation Theorem** was earlier stated and used in order to establish **Farkas' Lemma**. The **VI** stationarity conditions, and **Weierstrass' Theorem**, can be used to establish it.

> Suppose that $C \subseteq \mathbb{R}^n$ is closed and convex, and that the point $\mathbf{y}$ does not lie in C. Then there exist a vector $\boldsymbol{\pi} \neq \mathbf{0}^n$ and $\alpha \in \mathbb{R}$ such that $\boldsymbol{\pi}^{\mathrm{T}}\mathbf{y} > \alpha$ and $\boldsymbol{\pi}^{\mathrm{T}}\mathbf{x} \leq \alpha$ for all $\mathbf{x} \in C$



- *Proof*. Consider the problem of finding the closest point to $\mathbf{y}$ in C, that is,
$$\min_{\mathbf{x} \in S} h(\mathbf{x})$$
where $h(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2$
- Weierstrass guarantees an optimal solution $\mathbf{x}^* \in C$.
- (VI) $\Rightarrow \nabla h(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0$ for all $x \in C$.
- $\nabla h(\mathbf{x}) = \mathbf{x} - \mathbf{y}$
- Set $\boldsymbol{\pi} := \mathbf{y} - \mathbf{x}^* = -\nabla h(\mathbf{x}^*)$ and $\alpha := \boldsymbol{\pi}^{\mathrm{T}} \mathbf{x}^*$
- $\mathbf{y} \notin C$: $\boldsymbol{\pi}^{\mathrm{T}} \mathbf{y} - \alpha = (\mathbf{y} - \mathbf{x}^*)^{\mathrm{T}} \mathbf{y} - (\mathbf{y} - \mathbf{x}^*)^{\mathrm{T}} \mathbf{x}^* = \|\mathbf{y} - \mathbf{x}^*\|^2 > 0$
- For any $\mathbf{x} \in C$, the (VI) yields, $-\boldsymbol{\pi}^{\mathrm{T}}(\mathbf{x} - \mathbf{x}^*) \geq 0 \iff \boldsymbol{\pi}^{\mathrm{T}} \mathbf{x} \leq \alpha$
- The hyperplane provided by the theorem is actually a tangent to C, while the normal is $\mathbf{y} - \mathbf{x}^*$

- In the previous slide we used the solution to the problem $\min_{\mathbf{x}\in S} \|\mathbf{x} - \mathbf{y}\|^2$.
- As the objective function $\|\mathbf{x} - \mathbf{y}\|^2$ is strictly convex, the solution $\mathbf{x}^*$ is actually a unique solution.
- So for any $\mathbf{y} \in \mathbb{R}^n$, and a closed convex set S, there is always a **unique** nearest point $\mathbf{x}^* \in S$ to $\mathbf{y}$.
- We call this point the **projection** of $\mathbf{y}$ onto S. We denote this by $\mathrm{Proj}_S(\mathbf{y})$.
- Characterization from (VI): $\mathrm{Proj}_S(\mathbf{y})$ is the (unique) vector such that

$$(\mathbf{y} - \mathrm{Proj}_S(\mathbf{y}))^{\mathrm{T}}(\mathbf{x} - \mathrm{Proj}_S(\mathbf{y})) \leq 0, \quad \forall \mathbf{x} \in S$$

$\mathbf{x}^* \in S$ is stationary iff

$$\underset{\mathbf{x} \in S}{\text{minimum}} \; \nabla f(\mathbf{x}^*)^{\mathrm{T}}(\mathbf{x} - \mathbf{x}^*) = 0$$

- *Proof.*
- Method basis: given $\mathbf{x}_k \in S$, find out if we are stationary by minimizing $\nabla f(\mathbf{x}_k)^{\mathrm{T}}(\mathbf{x} - \mathbf{x}_k)$ over $\mathbf{x} \in S$. In some sense, we find the $\mathbf{x} \in S$ which "violates optimality the most." Perform a line search in the direction from $\mathbf{x}_k$ towards that point. Repeat until convergence
- Names: Frank–Wolfe, Simplicial decomposition. Chapter 12

$\mathbf{x}^* \in S$ is stationary iff

$$\mathbf{x}^* = \mathrm{Proj}_S[\mathbf{x}^* - \nabla f(\mathbf{x}^*)]$$

- *Proof.*
- In other words, $\mathbf{x}^*$ is stationary if and only if a step in the direction of the steepest descent direction followed by a Euclidean projection onto S means that we have not moved at all. (If not, then we obtain a descent direction towards that projected point—basis for the **projection method** in Chapter 12)

- The condition $\mathbf{x}^* = \mathrm{Proj}_S(\mathbf{x}^* - \nabla f(\mathbf{x}^*))$ holds if the angle between $-\nabla f(\mathbf{x}^*)$ and any feasible direction is at least 90°. These 'outwards' directions is called the **normal cone** at $\mathbf{x}^*$, denoted $N_S(\mathbf{x}^*)$.

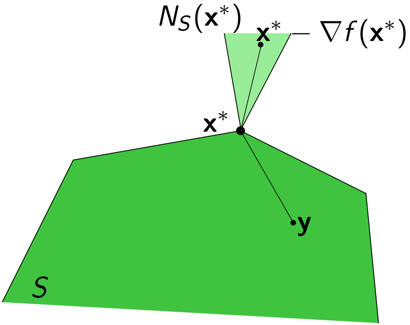


Figure: Normal cone characterization of a stationary point

- More formally, suppose $S \subseteq \mathbb{R}^n$ is closed and convex. Let $\mathbf{x} \in \mathbb{R}^n$. Then, the **normal cone** to S at $\mathbf{x}$ is the set

$$N_S(\mathbf{x}) := \begin{cases} \{\, \mathbf{v} \in \mathbb{R}^n \mid \mathbf{v}^{\mathrm{T}}(\mathbf{y} - \mathbf{x}) \leq 0, \quad \mathbf{y} \in S \,\}, & \text{if } \mathbf{x} \in S, \\ \emptyset & \text{otherwise} \end{cases}$$

Characterization of stationary point at $\mathbf{x}^* \in S$, number IV:

$$-\nabla f(\mathbf{x}^*) \in N_S(\mathbf{x}^*) \tag{3}$$

- Equivalent to $\mathbf{x}^* = \mathrm{Proj}_S(\mathbf{x}^*)$.
- If S is a subspace $\Longrightarrow \nabla f(\mathbf{x}^*)$ is a normal to the subspace!
- Note: if $\mathbf{x}^*$ interior point $\Longrightarrow N_S(\mathbf{x}^*) = \{\mathbf{0}^n\}$ ($S = \mathbb{R}^n \Longrightarrow \nabla f(\mathbf{x}^*) = \mathbf{0}^n$)
- Condition IV is the only version of the necessary conditions for convex sets that extends to non-convex sets (will be

- $\mathbf{x}^*$ local min on closed convex set $S \implies \mathbf{x}^*$ stationary
- $\mathbf{x}^*$ stationary AND problem convex $\implies \mathbf{x}^*$ global min on S
- $S = \mathbb{R}^n$: $\nabla f(\mathbf{x}^*) = \mathbf{0}^n$
- Four equivalent stationarity conditions for convex sets S:
  1. $S \subset \mathbb{R}^n$: (I) VI: $\nabla f(\mathbf{x}^*)^{\mathrm{T}}(\mathbf{x} - \mathbf{x}^*) \geq 0$ for all $\mathbf{x} \in S$
  2. $S \subset \mathbb{R}^n$: (II) Projection: $\mathbf{x}^* = \mathrm{Proj}_S[\mathbf{x}^* - \nabla f(\mathbf{x}^*)]$
  3. $S \subset \mathbb{R}^n$: (III) LP: $\min_{\mathbf{y} \in S} \nabla f(\mathbf{x}^*)^{\mathrm{T}}(\mathbf{y} - \mathbf{x}^*) = 0$
  4. $S \subset \mathbb{R}^n$: (IV) Normal cone: $-\nabla f(\mathbf{x}^*) \in N_S(\mathbf{x}^*)$
- Only (IV) can be extended to the case of non-convex sets S (the Karush–Kuhn–Tucker [KKT] conditions, Lecture 5–6)